Wyner's Common Information: Generalizations and A New Lossy Source Coding Interpretation

Ge Xu, Wei Liu, *Student Member, IEEE,* and Biao Chen, *Senior Member, IEEE*

Abstract

Wyner's common information was originally defined for a pair of dependent discrete random variables. Its significance is largely reflected in, hence also confined to, several existing interpretations in various source coding problems. This paper attempts to both generalize its definition and to expand its practical significance by providing a new operational interpretation. The generalization is two-folded: the number of dependent variables can be arbitrary, so are the alphabet of those random variables. New properties are determined for the generalized Wyner's common information of N dependent variables. More importantly, a lossy source coding interpretation of Wyner's common information is developed using the Gray-Wyner network. In particular, it is established that the common information equals to the smallest common message rate when the total rate is arbitrarily close to the rate distortion function with joint decoding. A surprising observation is that such equality holds independent of the values of distortion constraints as long as the distortions are within some distortion region. Examples about the computation of common information are given, including that of a pair of dependent Gaussian random variables.

Index Terms

Common information, Gray-Wyner network, rate distortion function

I. Introduction

Consider a pair of dependent random variables X and Y with joint distribution $p(x,y)$, which denotes either the probability density function if X and Y are continuous or the probability mass function if X and Y are discrete. Quantifying the information that is common between X and Y has been a classical problem both in information theory and in mathematical statistics [1]–[4]. The most widely used notion is Shannon's mutual information, defined as

$$I(X;Y) = E\left[\log \frac{p(x,y)}{p(x)p(y)}\right]$$

G. Xu and B. Chen are with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244 USA (email: gexu@syr.edu, bichen@syr.edu).

W. Liu was with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244 USA. He is now with Bloomberg L.P., New York, NY 10022 USA (email:wliusyr@gmail.com).

The material in this paper was presented in part at the Annual Allerton Conference on Communication, Control, and Computing, Montecillo, IL, Sept. 2010 and Annual Conference on Information Sciences and Systems, Baltimore, MD, March 2011. This work was supported in part by the Army Research Office under Award W911NF-12-1-0383, by the Air Force Office of Scientific Research under Award FA9550-10-1-0458, and by the National Science Foundation under Award 1218289.

where $p(x)$ and $p(y)$ are the marginal distribution of X and Y corresponding to the joint distribution $p(x, y)$ and $E[\cdot]$ denotes expectation with respect to $p(x, y)$. Shannon's mutual information measures the amount of uncertainty reduction in one variable by observing the other. Its significance lies in its applications to a broad range of problems in which concrete operational meanings of $I(X; Y)$ can be established. These include both source and channel coding problems in information and communication theory [5] and hypothesis testing problems in statistical inference [6].

Other notions of information have also been defined between a pair of dependent variables. Most notable among them are Gács and Körner's common randomness $K(X, Y)$ [2] and Wyner's common information $C(X, Y)$ [4]. Gács and Körner's common randomness is defined as the maximum number of common bits per symbol that can be independently extracted from X and Y. Quite naturally, $K(X, Y)$ has found extensive applications in secure communications, e.g., for key generation [7]–[9]. More recently, a new interpretation of $K(X, Y)$ using the Gray-Wyner source coding network was given in [10]. It was noted in [2], [11] that the definition of $K(X, Y)$ is rather restrictive in that $K(X, Y)$ equals 0 in most cases except for the special case when $X = (X', V)$ and $Y = (Y', V)$ and X', Y', V are independent variables or those (X, Y) pair that can be converted to such a dependence structure through relabeling the realizations, i.e., whose distribution is a permutation of the original joint distribution matrix. Notice also that $K(X, Y)$ is defined only for discrete random variables.

Wyner's common information was originally defined for a pair of discrete random variables with finite alphabet as

$$C(X, Y) = \inf_{X - W - Y} I(X, Y; W). \tag{1}$$

Here, the infimum is taken over all auxiliary random variables W such that X, W, and Y form a Markov chain. Clearly, the quantity $C(X, Y)$ in (1) can be defined for any pair of random variables with arbitrary alphabets. However, the operational meanings of $C(X, Y)$ available in existing literature are largely confined to that for discrete X and Y. These include the minimum common rate for the Gray-Wyner lossless source coding problem under a sum rate constraint, the minimum rate of a common input of two independent random channels for distribution approximation [4], and strong coordination capacity of a two-node network without common randomness and with actions assigned at one node [12].

This paper intends to generalize Wyner's common information along two directions. The first is to generalize it to that of multiple dependent random variables. The second is to generalize it to that of continuous random variables.

For the first direction, Wyner's common information is defined through a conditional independence structure which is equivalent to the Markov chain condition for two dependent variables. Relevant properties related to this generalization are derived. In addition, we prove that Wyner's original interpretations in [4] can be directly extended to that involving multiple variables. Note that both mutual information and common randomness have also been generalized to that of multiple random variables [14]–[16].

For the second direction, we provide a new lossy source coding interpretation using the Gray-Wyner network. Specifically, we show that, for the Gray-Wyner network, Wyner's common information is precisely the smallest common message rate for a certain range of distortion constraints when the total rate is arbitrarily close to the rate

distortion function with joint decoding. As the common information is only a function of the joint distribution, this smallest common rate remains constant even if the distortion constraints vary, as long as they are in a specific distortion region. There has also been recent effort in characterizing the common message rate for lossy source coding using the Gray-Wyner network [17]. We establish the equivalence between the characterization in [17] with an alternative characterization presented in the present paper.

Computing Wyner's common information is known to be a challenging problem; $C(X,Y)$ was only resolved for several special cases described in [4], [13]. Along with our generalizations of Wyner's common information, we provide two new examples where we can explicitly evaluate the common information of multiple dependent variables. In particular, we derive, through an estimation theoretic approach, $C(X,Y)$ for a bivariate Gaussian source and its extension to the multi-variate case with a certain correlation structure.

The rest of the paper is organized as follows. Section II reviews Wyner's two approaches for the common information of two discrete random variables, the general Gray-Wyner network, and the relations among joint, marginal, and conditional rate distortion functions. Section III gives the definition of Wyner's common information for N dependent random variables with arbitrary alphabets along with some associated properties. The operational meanings of Wyner's common information developed in [4] are also extended to that of N discrete dependent random variables in Section III. In Section IV, we provide a new interpretation of Wyner's common information using Gray-Wyner's lossy source coding network. Specifically, we prove that for the Gray-Wyner network, Wyner's common information is precisely the smallest common message rate for a certain range of distortion constraints when the total rate is arbitrarily close to the rate distortion function with joint decoding. In Section V, two examples, the doubly symmetric binary source and the bivariate Gaussian source, are used to illustrate the lossy source coding interpretation of Wyner's common information. The common information for bivariate Gaussian source and its extension to the multi-variate case is also derived in V. Section VI concludes this paper.

Notation: Throughout this paper, we use calligraphic letter $\mathcal{X}$ to denote the alphabet and $p(x)$ to denote either point mass function or probability density function of a random variable X. Boldface capital letter $\mathbf{X}^A$ denotes a vector of random variables $\{X_i\}_{i\in A}$ where A is an index set. $A\backslash B$ denotes set theoretic subtraction, i.e., $A\backslash B = \{x : x \in A \text{ and } x \notin B\}$. For two real vectors of identical size $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{x} \leq \mathbf{y}$ denotes component-wise inequality.

II. Existing Results

A. Wyner's result

Wyner defined the common information of two discrete random variables X and Y with distribution $p(x,y)$ in equation (1) and provided two operational meanings for this definition. The first approach is shown in Fig. 1. This model is a source coding network first studied by Gray and Wyner in [18]. In this model, the encoder observes a pair of sequences (X^n, Y^n), and map them to three messages W_0, W_1, W_2, taking values in alphabets of respective sizes $2^{nR_0}, 2^{nR_1}$ and 2^{nR_2}. Decoder 1, upon receiving (W_0, W_1), needs to reproduce X^n with high reliability while decoder 2, upon receiving (W_0, W_2), needs to reproduce Y^n with high reliability. Define

$$\Delta = \frac{1}{2n}\left(E[d_H(X^n,\hat{X}^n)] + E[d_H(Y^n,\hat{Y}^n)]\right)$$

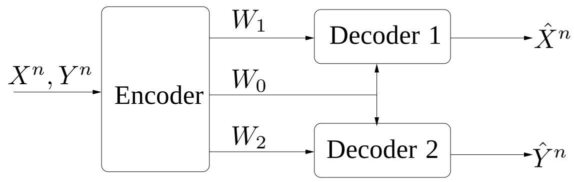

Fig. 1. Source coding over a simple network.

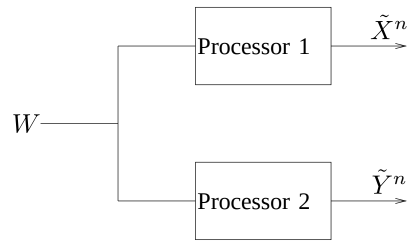

Fig. 2. Random variable generators.

where $d_H(\cdot,\cdot)$ is the Hamming distortion. Let C_1 be the the infimum of all achievable R_0 for the system in Fig. 1 such that for any $\epsilon > 0$, there exists, for n sufficiently large, a source code with the total rate $R_0+R_1+R_2 \le H(X,Y)+\epsilon$ and $\Delta \le \epsilon$.

The second approach is shown in Fig. 2. In this approach, the joint distribution $p(x^n, y^n) = \prod_{i=1}^n p(x_i, y_i)$ is approximated by the output distribution of a pair of random number generators. A common input W, uniformly distributed on $\mathcal{W} = \{1, \cdots, 2^{nR_0}\}$ is sent to two separate processors which are independent of each other. These processors (random number generators) generate independent and identically distributed (i.i.d) sequences according to two distributions $q_1(x^n|w)$ and $q_2(y^n|w)$ respectively. The output sequences of the two processors are denoted by $\tilde{X}^n$ and $\tilde{Y}^n$ respectively and the joint distribution of the output sequences is given by

$$q(x^n, y^n) = \sum_{w \in \mathcal{W}} \frac{1}{|\mathcal{W}|} q_1(x^n|w) q_2(y^n|w).$$

Let

$$D_n(q,p) = \frac{1}{n} \sum_{x^n \in \mathcal{X}^n, y^n \in \mathcal{Y}^n} q(x^n, y^n) \log \frac{q(x^n, y^n)}{p(x^n, y^n)}.$$

Let C_2 be the infimum of rate R_0 for the common input such that for any $\epsilon > 0$, there exists a pair of distribtions $q_1(x^n|w)$, $q_2(y^n|w)$ and n such that $D_n(q,p) \le \epsilon$.

Wyner proved in [4] that

$$C_1 = C_2 = C(X,Y).$$

B. Generalized Gray-Wyner networks

Consider the Gray-Wyner source coding network [18] with one encoder and N decoders as shown in Fig. 3. The encoder observes an i.i.d. vector source sequence $\{\mathbf{X}_1, \cdots, \mathbf{X}_n\}$ where each $\mathbf{X}_k = \{X_{1k}, \cdots, X_{Nk}\}$, $k = 1, \cdots, n$,

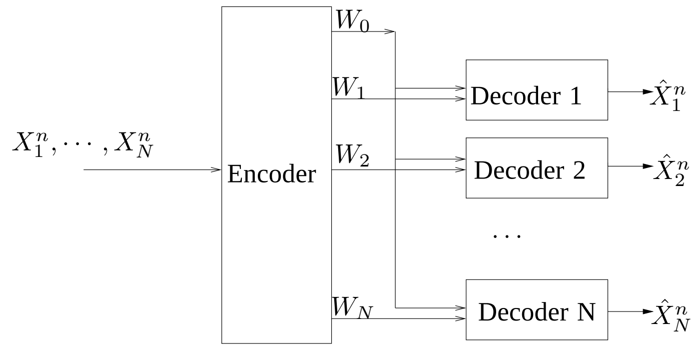

Fig. 3. Generalized Gray-Wyner source coding network.

is a length-N vector with joint distribution $p(\mathbf{x})$. Denote by $X_i^n = [X_{i1}, \cdots, X_{in}]$ the ith component of the vector sequence. There are a total of N receivers, with the ith receiver only interested in recovering the ith component sequence X_i^n. The encoder encodes the source into $N+1$ messages, one is a public message available at all receivers while the other N messages are private messages only available at the corresponding receivers.

For $m = 1, 2, \cdots$, let $I_m = \{0, 1, 2, \cdots, m-1\}$. An $(n, M_0, M_1, \cdots, M_N)$ code is defined by

- An encoder mapping

$$f : \mathcal{X}_1^n \times \cdots \times \mathcal{X}_N^n \to I_{M_0} \times I_{M_1} \times \cdots I_{M_N},$$

- N decoder mappings

$$g_i : I_{M_i} \times I_{M_0} \to \hat{\mathcal{X}}_i^n, \qquad i = 1, 2, \cdots, N.$$

For an $(n, M_0, M_1, \cdots, M_N)$ code, let $f(\mathbf{X}_1, \cdots, \mathbf{X}_n) = (W_0, W_1, \cdots, W_N)$ and $\hat{X}_i^n = g_i(W_i, W_0)$, $i = 1, 2, \cdots, N$.

We discuss below the lossless and lossy source coding using the generalized Gray-Wyner network.

1) Lossless Gray-Wyner source coding:

Define the probability of error as

$$P_e^{(n)} = \frac{1}{nN} \sum_{i=1}^{N} E[d_H(X_i^n, \hat{X}_i^n)], \tag{2}$$

where $\hat{X}_i^n = g_i(W_i, W_0)$ for $i = 1, \cdots, N$ and $d_H(u^n, \hat{u}^n)$ is the Hamming distance between u^n and $\hat{u}^n$.

A rate tuple $(R_0, R_1, \cdots, R_N)$ is said to be *achievable* if for any $\epsilon > 0$, there exists, for n sufficiently large, an $(n, M_0, M_1, \cdots, M_N)$ code such that

$$M_i \leq 2^{n(R_i+\epsilon)}, \qquad i = 0, 1, \cdots, N, \tag{3}$$

$$P_e^{(n)} \leq \epsilon. \tag{4}$$

Denote by $\mathcal{R}_1$ the region of all achievable rate tuples $(R_0, R_1, \cdots, R_N)$.

Theorem 1: $\mathcal{R}_1$ is the union of all rate tuples $(R_0, R_1, \cdots, R_N)$ that satisfy

$$R_0 \geq I(X_1, \cdots, X_N; W), \tag{5}$$

$$R_i \geq H(X_i|W), \quad i = 1, 2, \cdots, N, \tag{6}$$

for some $W \sim p(w|x_1, \cdots, x_N)$.

2) Lossy Gray-Wyner source coding:

Let $\mathbf{d}(\mathbf{x}, \hat{\mathbf{x}}) \triangleq \{d_1(x_1, \hat{x}_1), \cdots, d_N(x_N, \hat{x}_N)\}$ be a compound distortion measure. Define $\Delta_i, i = 1, \cdots, N$ to be the average distortion between the ith component sequence of the encoder input and the ith decoder output,

$$\Delta_i = E[d_i(X_i^n, \hat{X}_i^n)] = \frac{1}{n} \sum_{k=1}^{n} E[d_i(X_{ik}, \hat{X}_{ik})]. \tag{7}$$

Define the vector of average distortions to be $\mathbf{\Delta} \triangleq \{\Delta_1, \cdots, \Delta_N\}$. An $(n, M_0, M_1, \cdots, M_N)$ code with an average distortion vector $\mathbf{\Delta}$ is said to be an $(n, M_0, M_1, \cdots, M_N, \mathbf{\Delta})$ rate distortion code. Let $\mathbf{D} \triangleq \{D_1, D_2, \cdots, D_N\} \in \mathbb{R}_+^N$. A rate tuple $(R_0, R_1, \cdots, R_N)$ is said to be $\mathbf{D}$-*achievable* if for arbitrary $\epsilon > 0$, there exists, for n sufficiently large, an $(n, M_0, M_1, \cdots, M_N, \mathbf{\Delta})$ code such that

$$M_i \leq 2^{n(R_i+\epsilon)}, \quad i = 0, 1, \cdots, N, \tag{8}$$

$$\mathbf{\Delta} \leq \mathbf{D} + \epsilon. \tag{9}$$

Let $\mathcal{R}_2(\mathbf{D})$ be the region of all $\mathbf{D}$-achievable rate tuples $(R_0, R_1, \cdots, R_N)$.

Theorem 2: $\mathcal{R}_2(\mathbf{D})$ is the union of all rate tuples $(R_0, R_1, \cdots, R_N)$ that satisfy

$$R_0 \geq I(X_1, \cdots, X_N; W), \tag{10}$$

$$R_i \geq R_{X_i|W}(D_i), \quad i = 1, 2, \cdots, N, \tag{11}$$

for some $W \sim p(w|x_1, \cdots, x_N)$.

Here, $R_{X_i|W}(D_i)$ is the conditional rate distortion function defined as [21]

$$R_{X_i|W}(D_i) = \min_{p_t(\hat{x}_i|x_i,w): Ed_i(X_i,\hat{X}_i) \leq D_i} I(X_i; \hat{X}_i|W). \tag{12}$$

Theorems 1 and 2 are direct extensions of Theorem 4 and 8 in [18] for Gray-Wyner network with two receivers. Note that in [18], the authors proved only the discrete case for [18, Theorem 8], the proof for continuous alphabets can be constructed in a similar fashion.

C. Joint, marginal and conditional rate distortion functions

In this section, we review the joint, marginal and conditional rate distortion functions and their relations. Two-dimensional sources will be considered and the results can be generalized immediately to N-dimensional vector sources.

Given a two-dimensional source (X_1, X_2) with probability distribution $p(x_1, x_2)$ and two distortion measures $d_1(x_1, \hat{x}_1)$ and $d_2(x_2, \hat{x}_2)$ defined on $\mathcal{X}_1 \times \hat{\mathcal{X}}_1$ and $\mathcal{X}_2 \times \hat{\mathcal{X}}_2$, the joint rate distortion function is given by

$$R_{X_1X_2}(D_1, D_2) = \min I(X_1X_2; \hat{X}_1\hat{X}_2), \tag{13}$$

where the minimum is taken over all test channels $p_t(\hat{x}_1\hat{x}_2|x_1x_2)$ such that $Ed_1(X_1,\hat{X}_1)\le D_1$, $Ed_2(X_2,\hat{X}_2)\le D_2$. The conditional rate distortion function is defined in (12). The joint, marginal and conditional rate distortion functions satisfy the following inequalities.

Lemma 1: [19], [20] Given a two-dimensional source (X_1,X_2) with joint distribution $p(x_1,x_2)$ and two distortion measures $d_1(x_1,\hat{x}_1)$, $d_2(x_2,\hat{x}_2)$ defined respectively on $\mathcal{X}_1\times\hat{\mathcal{X}}_1$ and $\mathcal{X}_2\times\hat{\mathcal{X}}_2$, the rate distortion functions satisfy the following inequalities

$$R_{X_1X_2}(D_1,D_2)\ge R_{X_1|X_2}(D_1)+R_{X_2}(D_2), \tag{14a}$$

$$R_{X_1|X_2}(D_1)\ge R_{X_1}(D_1)-I(X_1;X_2), \tag{14b}$$

$$R_{X_1X_2}(D_1,D_2)\ge R_{X_1}(D_1)+R_{X_1}(D_2)-I(X_1;X_2), \tag{14c}$$

$$R_{X_1}(D_1)\ge R_{X_1|X_2}(D_1), \tag{15a}$$

$$R_{X_1}(D_1)+R_{X_2}(D_2)\ge R_{X_1X_2}(D_1,D_2). \tag{15b}$$

Sufficient conditions for equality in (14) are that the optimum backward test channels for the functions on the left side of each equation factor appropriately, i.e., for (14a) $p_b(x_1x_2|\hat{x}_1\hat{x}_2)=p(x_1|\hat{x}_1x_2)p(x_2|\hat{x}_2)$, for (14b) $p_b(x_1|\hat{x}_1x_2)=p(x_1|\hat{x}_1)$ and for (14c) that $p_b(x_1x_2|\hat{x}_1\hat{x}_2)=p(x_1|\hat{x}_1)p(x_2|\hat{x}_2)$. Equalities hold in (15) if and only if X_1 and X_2 are independent.

Furthermore, Gray has shown that under quite general conditions, equalities hold in (14) for small values of distortion. This is because the marginal, joint and conditional rate distortion functions equal to their Extended Shannon Lower Bounds (ESLB) [19], [21] under suitable conditions. These ESLB, denoted by $R_X^{(L)}(D)$ for a rate distortion function $R_X(D)$, satisfy the following property. Denote by $\mathcal{D}$ a surface in the m-dimensional space and the inequality $\Delta\le\mathcal{D}$ means that there exists a vector $\beta\in\mathcal{D}$ such that $\Delta\le\beta$. If there is no such a vector, $\Delta>\mathcal{D}$. Likewise, $\mathcal{D}_1\le\mathcal{D}_2$ means that $\beta\le\mathcal{D}_2$ for any $\beta\in\mathcal{D}_1$ [19].

Lemma 2: [19] Given a two-dimensional source (X_1,X_2) with joint distribution $p(x_1,x_2)$ such that for $x_1\in\mathcal{X}_1, x_2\in\mathcal{X}_2$, $p(x_2|x_1)>0$, reproduction alphabets $\hat{\mathcal{X}}_1=\mathcal{X}_1$, $\hat{\mathcal{X}}_2=\mathcal{X}_2$ and two per-letter distortion measures $d_1(x_1,\hat{x}_1)$ and $d_2(x_2,\hat{x}_2)$ that satisfy

$$d_i(x_i,\hat{x}_i)\quad>\quad d_i(x_i,x_i)=0, x_i\ne\hat{x}_i, i=1,2, \tag{16}$$

there exist strictly positive surfaces $\mathcal{D}(X_1X_2)$, $\mathcal{D}(X_1|X_2)$, $\mathcal{D}(X_1)$ and $\mathcal{D}(X_2)$ such that

$$\begin{array}{ll}
R_{X_1X_2}(D_1,D_2)=R^{(L)}_{X_1X_2}(D_1,D_2), & \text{if } (D_1,D_2)\le\mathcal{D}(X_1X_2),\\
R_{X_1|X_2}(D_1)=R^{(L)}_{X_1|X_2}(D_1), & \text{if } D_1\le\mathcal{D}(X_1|X_2),\\
R_{X_1}(D_1)=R^{(L)}_{X_1}(D_1), & \text{if } D_1\le\mathcal{D}(X_1),\\
R_{X_2}(D_2)=R^{(L)}_{X_2}(D_2), & \text{if } D_2\le\mathcal{D}(X_2),
\end{array}$$

and

$$\begin{aligned}
\mathcal{D}(X_1|X_2)\quad&\le\quad\mathcal{D}(X_1),\\
\mathcal{D}(X_1X_2)\quad&\le\quad(\mathcal{D}(X_1|X_2),\mathcal{D}(X_2))\le(\mathcal{D}(X_1),\mathcal{D}(X_2))\,.
\end{aligned}$$

Finally,

$$R^{(L)}_{X_1X_2}(D_1,D_2) = R^{(L)}_{X_1|X_2}(D_1)+R^{(L)}_{X_2}(D_2), \tag{17}$$

$$= R^{(L)}_{X_1}(D_1)+R^{(L)}_{X_2}(D_2)-I(X_1;X_2). \tag{18}$$

It is apparent that when the rate distortion functions equal to their corresponding ESLB, equations (17) and (18) imply equalities in (14a), (14b) and (14c).

III. The Common Information of N Dependent Discrete Random Variables

A. Definition

Wyner's original definition of the common information in (1) assumes a Markov chain between the random variables X, Y and the auxiliary variable W, i.e., $X-W-Y$. This Markov chain is equivalent to stating that X and Y are conditionally independent given W. This conditional independence structure can be naturally generalized to that of N dependent random variables. Let $\mathbf{X}\triangleq\{X_1,\cdots,X_N\}$ be N dependent random variables that take values in some arbitrary (finite, countable, or continuous) spaces $\mathcal{X}_1\times\mathcal{X}_2\times\cdots\times\mathcal{X}_N$. The joint distribution of $\mathbf{X}$ is denoted as $p(\mathbf{x})$, which is either a probability mass function or a probability density function. We now give the definition of the common information for N dependent random variables.

Definition 1: Let $\mathbf{X}$ be a random vector with joint distribution $p(\mathbf{x})$. The common information of $\mathbf{X}$ is defined as

$$C(\mathbf{X})\triangleq\inf I(\mathbf{X};W), \tag{19}$$

where the infimum is taken over all the joint distributions of $(\mathbf{X},W)$ such that

1) the marginal distribution for $\mathbf{X}$ is $p(\mathbf{x})$,
2) $\mathbf{X}$ are conditionally independent given W, i.e.,

$$p(\mathbf{x}|w) = \prod_{i=1}^{N}p(x_i|w). \tag{20}$$

We now discuss several properties associated with the definition given in (19).

Wyner's common information of two random variables (X_1,X_2) satisfies the following inequality

$$I(X_1,X_2)\le C(X_1,X_2)\le\min\{H(X_1),H(X_2)\}.$$

A similar inequality for the common information of N random variables can be derived. Let $A\subseteq\mathcal{N}=\{1,2,\cdots,N\}$ and $\bar{A}=\mathcal{N}\backslash A$. We have

$$\max_A\{I(\mathbf{X}^A;\mathbf{X}^{\bar{A}})\}\le C(\mathbf{X})\le\min_j\{H(\mathbf{X}^{-j})\}, \tag{21}$$

where $\mathbf{X}^{-j}\triangleq\mathbf{X}^{\mathcal{N}\backslash\{j\}}=\{X_1,\cdots,X_{j-1},X_{j+1},\cdots,X_N\}$ for $j\in\mathcal{N}$.

To verify the upper bound, for any $j\in\mathcal{N}$, let $W_j=\mathbf{X}^{-j}$. Thus, $X_1,\cdots,X_N$ are conditionally independent given W_j, and

$$I(\mathbf{X};W_j)=I(\mathbf{X};\mathbf{X}^{-j})=H(\mathbf{X}^{-j}).$$

Thus $C(\mathbf{X}) \leq H(\mathbf{X}^{-j})$ for all $j \in \mathcal{N}$.

For the lower bound, since $X_1, \cdots, X_N$ are conditionally independent given W, we have the Markov chain $\mathbf{X}^A - W - \mathbf{X}^{\bar{A}}$ for any subset $A \subseteq \mathcal{N}$. Hence,

$$I(\mathbf{X}; W) \geq I(\mathbf{X}^A; W) \geq I(\mathbf{X}^A; \mathbf{X}^{\bar{A}}),$$

where the second inequality is by the data processing inequality.

Therefore,

$$I(\mathbf{X}; W) \geq \max_A \{I(\mathbf{X}^A; \mathbf{X}^{\bar{A}})\}. \tag{22}$$

The common information defined in (19) also satisfies the following monotone property.

Lemma 3: Let $\mathbf{X} \sim p(\mathbf{x})$. For any two sets A, B that satisfy $A \subseteq B \subseteq \mathcal{N} = \{1, 2, \cdots, N\}$, we have

$$C(\mathbf{X}^A) \leq C(\mathbf{X}^B), \tag{23}$$

Proof: Let W' be the auxiliary variable that achieves $C(\mathbf{X}^B)$, i.e., $I(\mathbf{X}^B; W') = \inf_W I(\mathbf{X}^B; W)$. Since $A \subseteq B$, $\mathbf{X}^B$ being conditionally independent given W' implies that $\mathbf{X}^A$ are conditionally independent given W'. Thus

$$\begin{aligned} I(\mathbf{X}^B; W') &\geq I(\mathbf{X}^A; W'), \\ &\geq \inf I(\mathbf{X}^A; W), \end{aligned}$$

where the infimum is taken over all W such that $\mathbf{X}^A$ is independent given W. ■

The above monotone property of the common information is contrary to what the name implies: conceptually, the information in common ought to decrease when new variables are included in the set of random variables. Such is the case for Gács and Körner's common randomness, i.e., $K(\mathbf{X}^A) \geq K(\mathbf{X}^B)$. As a consequence, we have that for any N random variables $C(\mathbf{X}) \geq K(\mathbf{X})$. The fact that the common information $C(\mathbf{X})$ increases as more variables are involved suggests that it may have potential applications in statistical inference problems. This was explored in [22].

B. Coding theorems for the common information of N discrete random variables

Section II-A describes two operational interpretations of Wyner's common information for two discrete random variables based on the Gray-Wyner network and distribution approximation. These operational interpretations can also be extended to the common information of N dependent random variables.

For the first approach, we consider the lossless Gray-Wyner network with N terminals. For the Gray-Wyner source coding network, A number R_0 is said to be *achievable* if for any $\epsilon > 0$, there exists, for n sufficiently large, an $(n, M_0, M_1, \cdots, M_N)$ code with

$$M_0 \leq 2^{nR_0}, \tag{24}$$

$$\frac{1}{n}\sum_{i=0}^{N} \log M_i \leq H(\mathbf{X}) + \epsilon, \tag{25}$$

$$P_e^{(n)} \leq \epsilon. \tag{26}$$

Define C_1 as the infimum of all achievable R_0.

Theorem 3: For N discrete random variables $\mathbf{X}$ with joint distribution $p(\mathbf{x})$,

$$C_1 = C(\mathbf{X}). \tag{27}$$

The proof of Theorem 3 is a direct extension of the proof for two discrete random variables in [4] and hence is omitted.

The second approach of interpreting the common information of discrete random variable uses distribution approximation. Let $\{\mathbf{X}_1, \cdots, \mathbf{X}_n\}$ be i.i.d. copies of $\mathbf{X}$ with distribution $p(\mathbf{x})$, i.e., the joint distribution for $\{\mathbf{X}_1, \cdots, \mathbf{X}_n\}$ is

$$p^{(n)}(\mathbf{x}_1, \cdots, \mathbf{x}_n) = \prod_{k=1}^{n} p(\mathbf{x}_k). \tag{28}$$

An (n, M, Δ) *generator* consists of the following:

- a message set $\mathcal{W}$ with cardinality M;
- for all $w \in \mathcal{W}$, probability distributions $q_i^{(n)}(x_i^n|w)$, for $i = 1, 2, \cdots, N$.

Define the probability distribution on $\mathcal{X}_1^n \times \mathcal{X}_2^n \times \cdots \times \mathcal{X}_N^n$

$$q^{(n)}(\mathbf{x}_1, \cdots, \mathbf{x}_n) = \sum_{w \in \mathcal{W}} \frac{1}{M} \prod_{i=1}^{N} q_i^{(n)}(x_i^n|w). \tag{29}$$

Let

$$\Delta = D_n\left(q^{(n)}(\mathbf{x}_1, \cdots, \mathbf{x}_n); p^{(n)}(\mathbf{x}_1, \cdots, \mathbf{x}_n)\right) = \sum_{\mathbf{x}^n} \frac{1}{n} q^{(n)}(\mathbf{x}_1, \cdots, \mathbf{x}_n) \log \frac{q^{(n)}(\mathbf{x}_1, \cdots, \mathbf{x}_n)}{p^{(n)}(\mathbf{x}_1, \cdots, \mathbf{x}_n)}, \tag{30}$$

where $p^{(n)}(\mathbf{x}_1, \cdots, \mathbf{x}_n)$ is defined in (28) and $q^{(n)}(\mathbf{x}_1, \cdots, \mathbf{x}_n)$ is defined as in (29).

A number R is said to be *achievable* if for all $\epsilon > 0$, if for n sufficiently large there exists an (n, M, Δ) generator with $M \leq 2^{nR}$ and $\Delta \leq \epsilon$. Define C_2 as the infimum of all achievable R.

Theorem 4: For N discrete random variables $\mathbf{X}$ with joint distribution $p(\mathbf{x})$,

$$C_2 = C(\mathbf{X}). \tag{31}$$

The proof can be constructed in the same way as that of [4, Theorems 5.2 and 6.2], hence is omitted.

IV. The Lossy Source Coding Interpretation of Wyner's Common Information

The common information defined in (1) and (19) equally applies to that of continuous random variables. However, such definitions are only meaningful when they are associated with concrete operational interpretations. In this section, we develop a lossy source coding interpretation of Wyner's common information using the Gray-Wyner network. While this new interpretation holds for the general case of N dependent random variable, we elect to present coding theorems involving only a pair of dependent variables for ease of notion and presentation.

A. Lossy Gray-Wyner source coding

Given a two-dimensional source $(X_1, X_2) \sim p(x_1, x_2)$, for any $(D_1, D_2) \geq 0$, a number R_0 is said to be (D_1, D_2)-*achievable* if for any $\epsilon > 0$, there exists, for n sufficiently large, an $(n, M_0, M_1, M_2, \Delta_1, \Delta_2)$ code with

$$\begin{aligned} M_0 &\leq 2^{nR_0}, & (32)\\ \sum_{i=0}^{2} \frac{1}{n} \log M_i &\leq R_{X_1X_2}(D_1, D_2) + \epsilon, & (33)\\ \Delta_1 \leq D_1 + \epsilon \quad &, \quad \Delta_2 \leq D_2 + \epsilon. & (34) \end{aligned}$$

Define $C_3(D_1, D_2)$ as the infimum of all R_0's that are (D_1, D_2)-achievable. Thus, $C_3(D_1, D_2)$ is the minimum common message rate for the Gray-Wyner network with sum rate $R_{X_1X_2}(D_1, D_2)$ while satisfying the distortion constraint. Since $R_{X_1X_2}(D_1, D_2)$ is always (D_1, D_2)-achievable, it is obvious that

$$C_3(D_1, D_2) \leq R_{X_1X_2}(D_1, D_2). \tag{35}$$

The following theorem gives a precise characterization of $C_3(D_1, D_2)$.

Theorem 5:

$$C_3(D_1, D_2) = \tilde{C}(D_1, D_2), \tag{36}$$

where $\tilde{C}(D_1, D_2)$ is the solution of the following optimization problem:

$$\begin{aligned} &\inf && I(X_1, X_2; W) & (37)\\ &\text{subject to} && R_{X_1|W}(D_1) + R_{X_2|W}(D_2) + I(X_1, X_2; W) = R_{X_1X_2}(D_1, D_2). \end{aligned}$$

Proof: See Appendix A. ∎

The authors in [17] gave an alternative characterization of $C_3(D_1, D_2)$. Define

$$C^*(D_1, D_2) = \inf I(X_1, X_2; W),$$

where the infimum is taken over all joint distributions for $X_1, X_2, X_1^*, X_2^*, W$ such that

$$X_1^* - W - X_2^*, \tag{38}$$

$$(X_1, X_2) - (X_1^*, X_2^*) - W, \tag{39}$$

where (X_1^*, X_2^*) achieves $R_{X_1X_2}(D_1, D_2)$. It was shown in [17] that $C_3(D_1, D_2) = C^*(D_1, D_2)$. This, combined with Theorem 5, establishes that

$$\tilde{C}(D_1, D_2) = C^*(D_1, D_2). \tag{40}$$

$\tilde{C}(D_1, D_2)$ is derived from the rate distortion region $\mathcal{R}_2(D_1, D_2)$ given in Theorem 2 while the authors in [17] chose to derive $C^*(D_1, D_2)$ from an alternative characterization of $\mathcal{R}_2(D_1, D_2)$ given in [23]. In Appendix B, we provide a direct proof of (40) for completeness. Also, as given in Appendix B, a necessary condition for the equality condition in the optimization problem (37) is

$$R_{X_1X_2|W}(D_1, D_2) = R_{X_1|W}(D_1) + R_{X_2|W}(D_2).$$

B. The relation of $C_3(D_1,D_2)$ and the common information

Given our characterization of $C_3(D_1,D_2)$ in Theorem 5, we now establish its connection with $C(X_1,X_2)$ which leads to a new interpretation of Wyner's common information. We begin with the following two lemmas.

Lemma 4: Let W be the random variable that achieves the common information of X_1 and X_2. If

$$R_{X_1X_2|W}(D_1,D_2)+C(X_1,X_2)=R_{X_1X_2}(D_1,D_2),$$

then

$$C_3(D_1,D_2)\leq C(X_1,X_2). \tag{41}$$

Lemma 4 is a direct consequence of Theorem 5 as the Markov chain X_1-W-X_2 implies $R_{X_1X_2|W}(D_1,D_2)=R_{X_1|W}(D_1)+R_{X_2|W}(D_2)$. Thus, the equality constraint in (37) is satisfied. Inequality (41) follows as

$$\tilde{C}(D_1,D_2)=C_3(D_1,D_2)\leq I(X_1,X_2;W)=C(X_1,X_2).$$

The next lemma gives a sufficient condition under which $C_3(D_1,D_2)\geq C(X_1,X_2)$ is true.

Lemma 5: For any distortion pair (D_1,D_2), if the rate distortion function satisfies

$$R_{X_1X_2}(D_1,D_2)=R_{X_1}(D_1)+R_{X_2}(D_1)-I(X_1;X_2), \tag{42}$$

then we have

$$C_3(D_1,D_2)\geq C(X_1,X_2).$$

Proof: See Appendix C. ■

Lemmas 4 and 5, together with the relations of marginal, joint and conditional rate distortion functions described in Lemmas 1 and 2, allow us to determine a region such that $C_3(D_1,D_2)$ equals to the common information.

Theorem 6: Let random variables X_1,X_2 be distributed as $p(x_1,x_2)$ on $\mathcal{X}_1\times\mathcal{X}_2$ such that for $x_1\in\mathcal{X}_1, x_2\in\mathcal{X}_2$, $p(x_2|x_1)>0$. Let the reproduction alphabets $\hat{\mathcal{X}}_1=\mathcal{X}_1$, $\hat{\mathcal{X}}_2=\mathcal{X}_2$. The two per-letter distortion measures $d_1(x_1,\hat{x}_1)$, $d_2(x_2,\hat{x}_2)$ satisfy

$$d_i(x_i,\hat{x}_i) \quad > \quad d_i(x_i,x_i)=0, \quad x_i\neq\hat{x}_i, i=1,2. \tag{43}$$

Then there exists a strictly positive surface $\gamma\triangleq(\gamma_1,\gamma_2)$ such that, for $(D_1,D_2)\leq\gamma$,

$$C_3(D_1,D_2)=C(X_1,X_2). \tag{44}$$

Proof: See Appendix D. ■

Theorem 6 shows that Wyner's common information is precisely the smallest common message rate $C_3(D_1,D_2)$ of Gray-Wyner network for a certain range of distortion constraints when the total rate is arbitrarily close to the rate distortion function with joint decoding. As the common information is only a function of the joint distribution, hence is a constant for a given $p(x_1,x_2)$, it is surprising that the smallest common rate $C_3(D_1,D_2)$ remains constant even if the distortion constraints vary, as long as they are in a specific distortion region.

While Theorem 6 establishes that $C_3(D_1,D_2)=C(X_1,X_2)$ for $(D_1,D_2)\leq\gamma$, it does not specify the value of the positive distortion vector γ. Let $\mathcal{D}^c \triangleq (D_1^c,D_2^c)$ be the two-dimensional distortion surface such that

$R_{X_1X_2}(D_1^c, D_2^c) = C(X_1, X_2)$, then we must have that $\gamma \leq \mathcal{D}^c$. This is because if $\gamma > \mathcal{D}^c$, then there exists (D_1, D_2) such that $\gamma \geq (D_1, D_2) > \mathcal{D}^c$ and $C_3(D_1, D_2) \leq R_{X_1X_2}(D_1, D_2) < R_{X_1X_2}(D_1^c, D_2^c) = C(X_1, X_2)$, which contradicts Theorem 6. Now let us consider a particular point on the surface $\mathcal{D}^c$. Let W be the auxiliary random variable that achieves $C(X_1, X_2)$. Suppose there exists a distortion pair (D_1^0, D_2^0) satisfying, for $i = 1, 2$,

$$\begin{aligned} R_{X_i}(D_i^0) &= I(X_i; W), \\ D_i^0 &= \inf_{\hat{x}_i(w)} Ed_i(X_i, \hat{X}_i^0(W)), \end{aligned} \tag{45}$$

where $\hat{x}_1^0(w), \hat{x}_2^0(w)$ are deterministic functions. Under this assumption, we can show that $R_{X_1X_2}(D_1^0, D_2^0) = I(X_1, X_2; W)$. Therefore, the joint rate distortion function $R_{X_1X_2}(D_1^0, D_2^0)$ not only equals to the common information but also is achieved by the auxiliary random variable W. Furthermore, it is easy to show

$$C_3(D_1^0, D_2^0) = C(X_1, X_2), \tag{46}$$

using Lemma 5 and the fact that $C_3(D_1^0, D_2^0) \leq R_{X_1X_2}(D_1^0, D_2^0)$. This means that in the Gray-Wyner network, with the total rate equal to $R_{X_1X_2}(D_1^0, D_2^0)$, the scheme to transmit the pair of sources (X_1^n, X_2^n) within distortion constraints (D_1^0, D_2^0) is to communicate W to the two receivers using the common channel.

Let us now decrease the distortion constraints from (D_1^0, D_2^0) to $(D_1, D_2) \leq (D_1^0, D_2^0)$. The question is whether the rate $C(X_1, X_2)$ is $(D_1, D_2)-$achieveble, i.e., if it is possible to transmit the sources (X_1^n, X_2^n) with smaller distortions (D_1, D_2) with the sum rate at $R_{X_1X_2}(D_1, D_2)$ while keeping the common rate at $C(X_1, X_2)$. In the following, we identify a sufficient condition for $C_3(D_1, D_2) = C(X_1, X_2)$ for successively refinable sources. A source X with distortion measure $d(x, \hat{x})$ is said to be successively refinable from a coarser distortion δ_1 to a finer distortion δ_2 ($\delta_1 \geq \delta_2$) if it can be encoded in two stages in which the optimal descriptions at the second stage is a refinement of the optimal descriptions at the first stage [27]. Similar definition can be applied to vector sources with individual distortion constraints and the details can be found in [30].

In the following theorem, we give a sufficient condition under which $C_3(D_1, D_2) = C(X_1, X_2)$ for any $(D_1, D_2) \leq (D_1^0, D_2^0)$. This sufficient condition ensures the optimality of a two-stage encoding scheme: first encode the common message with rate $C(X_1, X_2)$ and we can obtain a coarse distortion (D_1^0, D_2^0), then encode the two private messages with rates $R_{X_1|W}(D_1)$ and $R_{X_2|W}(D_2)$. The successive refinement assumption guarantees that the two-step approach can achieve the distortion (D_1, D_2) and the sum rate does not exceed the total rate $R_{X_1X_2}(D_1, D_2)$.

Theorem 7: Let W be the auxiliary variable that achieves $C(X_1, X_2)$ and (D_1^0, D_2^0) be a distortion pair satisfying (45). If the source (X_1, X_2) is successively refinable from (D_1^0, D_2^0) to (D_1, D_2) for $(D_1, D_2) \leq (D_1^0, D_2^0)$, and X_i is successively refinable from D_i^0 to D_i for $D_i \leq D_i^0$, $i = 1, 2$, then,

$$C_3(D_1, D_2) = C(X_1, X_2).$$

Proof: See Appendix E. ■

In the following section, we will consider two examples involving successively refinable sources: the binary random variables and bivariate Gaussian variables. For these two cases, we compute explicitly the function $C_3(D_1, D_2)$

and establish its connection with $C(X_1, X_2)$. The distortion pair (D_1^0, D_2^0) satisfying (45) are identified for both cases, thus Theorem 7 can be directly applied.

V. EXAMPLES

A. Binary random variables

Let $S \sim \text{Bern}(\theta)$ for $0 \leq \theta \leq 1$, i.e., $S \in \{0,1\}$ and $P(S = 1) = \theta$. Let X_i, $i = 1, \cdots, N$, be the output of a binary symmetric channel (BSC) with crossover probability a_1 $(0 \leq a_1 \leq \frac{1}{2})$ and with S as input. The BSC channels are independent of each other. Thus,

$$p(x_1, \cdots, x_N|s) = \prod_{i=1}^{N} p(x_i|s),$$

where

$$p(x_i|s) = \begin{cases} 1 - a_1, & \text{if } x_i = s, \\ a_1, & \text{otherwise}, \end{cases}$$

for $x_i \in \{0, 1\}$. Therefore, the joint distribution of $X_1, X_2, \cdots, X_N$ is

$$\begin{aligned} p(x_1, x_2, \cdots, x_N) &= \sum_{s \in \{0,1\}} p(s) \prod_{i=1}^{N} p(x_i|s), \\ &= \theta a_1^{t_N} (1 - a_1)^{N - t_N} + (1 - \theta)(1 - a_1)^{t_N} a_1^{N - t_N}, \end{aligned} \tag{47}$$

where $t_N = \sum_{i=1}^{N} x_i$.

For $N = 2$, the joint distribution of X_1, X_2 is given by the following probability matrix,

$$\begin{bmatrix} \theta(1 - a_1)^2 + (1 - \theta)a_1^2 & a_1(1 - a_1) \\ a_1(1 - a_1) & \theta a_1^2 + (1 - \theta)(1 - a_1)^2 \end{bmatrix}. \tag{48}$$

It has been shown by Witsenhausen [13] that the common information of X_1, X_2 is achieved with W being S. That is

$$C(X_1, X_2) = I(X_1 X_2; S) = H(X_1, X_2) - 2h(a_1), \tag{49}$$

where $h(\cdot)$ is the binary entropy function. When $\theta = \frac{1}{2}$, (X_1, X_2) is a Doubly Symmetric Binary Source (DSBS) whose common information was derived by Wyner [4] using a different approach.

We now obtain the common information for N variables.

Proposition 1: Let $S \sim \text{Bern}(\theta)$ and let X_i, $i = 1, \cdots, N$, be the output of independent BSCs with common input S and crossover probability $0 \leq a_1 \leq 1/2$. Then for any $N \geq 2$, the common information for $X_1, \cdots, X_N$ is given as

$$C(X_1, \cdots, X_N) = I(X_1, \cdots, X_N; S). \tag{50}$$

Proof: That $C(X_1, \cdots, X_N) \le I(X_1, \cdots, X_N; S)$ follows from the definition of the common information (19). The inequality $C(X_1, \cdots, X_N) \ge I(X_1, \cdots, X_N; S)$ can be proved by contradiction. Suppose there exists a W such that

$$C(X_1, \cdots, X_N) = I(X_1, \cdots, X_N; W) < I(X_1, \cdots, X_N; S), \tag{51}$$

i.e., $C(X_1, \cdots, X_N)$ is achieved by W and it is strictly less than $I(X_1, \cdots, X_N; S)$. Since W induces conditional independence of $X_1, \cdots, X_N$, we have, from (51),

$$\sum_{i=1}^{N} H(X_i|W) > \sum_{i=1}^{N} H(X_i|S).$$

Thus, there must exist two random variables X_k, X_j, $k, j \in \{1, \cdots, N\}$ such that

$$H(X_k|W) + H(X_j|W) > H(X_k|S) + H(X_j|S).$$

Given that the sequence $\{X_1, \cdots, X_N\}$ is exchangeable [31], $p(x_k, x_j)$ has the same joint distribution as $p(x_1, x_2)$. Thus,

$$C(X_1, X_2) = C(X_k, X_j) = I(X_k, X_j; W) < I(X_k, X_j; S) = I(X_1, X_2; S).$$

This, however, contradicts the fact that S achieves $C(X_1, X_2)$. Thus the proposition is proved. ∎

We now characterize the minimum common rate $C_3(D_1, D_2)$ for a DSBS.

Proposition 2: Consider a DSBS (X_1, X_2) with distribution

$$p(x_1, x_2) = \begin{cases} \frac{1}{2}(1 - a_0), & \text{if } x_1 = x_2, \\ \frac{1}{2}a_0, & \text{otherwise}, \end{cases} \tag{52}$$

where, without loss of generality, $0 \le a_0 \le 1/2$. Let a_1 be such that $a_0 = 2a_1(1 - a_1), 0 \le a_1 \le 1/2$. With Hamming distortion $d_1 = d_2 = d_H$, we have

$$C_3(D_1, D_2) = \begin{cases} C(X_1, X_2), & (D_1, D_2) \in \mathcal{E}_{10}, \\ R_{X_1X_2}(D_1, D_2), & (D_1, D_2) \in \mathcal{E}_2 \cup \mathcal{E}_3, \\ 0, & (D_1, D_2) \ge (\frac{1}{2}, \frac{1}{2}), \end{cases} \tag{53}$$

$$C(X_1, X_2) \le C_3(D_1, D_2) \le R_{X_1X_2}(D_1, D_2), \quad (D_1, D_2) \in \mathcal{E}_{11}, \tag{54}$$

where

$$\begin{aligned} \mathcal{E}_{10} &= \{(D_1, D_2) : 0 \le D_i \le a_1, i = 1, 2\}, \\ \mathcal{E}_{11} &= \mathcal{E}_{10}^c \cap \{(D_1, D_2) : D_1 + D_2 - 2D_1D_2 \le a_0\}, \\ \mathcal{E}_2 &= \mathcal{E}_{10}^c \cap \mathcal{E}_{11}^c \cap \left\{(D_1, D_2) : \max\left\{\frac{D_1 - D_2}{1 - 2D_2}, \frac{D_2 - D_1}{1 - 2D_1}\right\} \le a_0\right\}, \\ \mathcal{E}_3 &= \mathcal{E}_{10}^c \cap \mathcal{E}_{11}^c \cap \mathcal{E}_2^c \cap \left\{(D_1, D_2) : D_i \le \frac{1}{2}, i = 1, 2\right\}. \end{aligned} \tag{55}$$

Proof: For $X_i \sim \text{Bern}(1/2), i = 1, 2$ with Hamming distortion, the rate distortion function is

$$R_{X_i}(D_i) = \begin{cases} 1 - h(D_i), & 0 \le D_i \le \frac{1}{2}, \\ 0, & D_i \ge \frac{1}{2}. \end{cases}$$

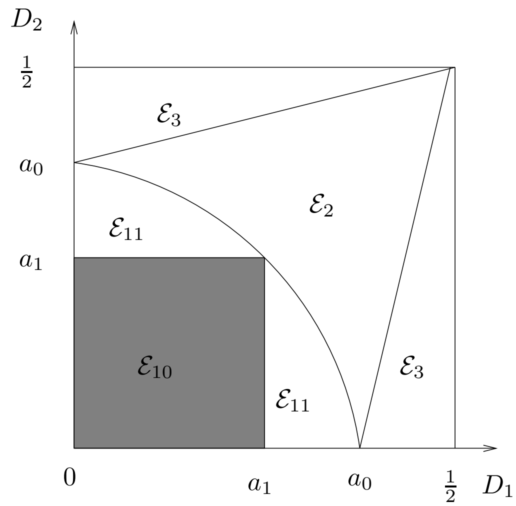

Fig. 4. The distortion regions $\mathcal{E}_{10}, \mathcal{E}_{11}, \mathcal{E}_2$ and $\mathcal{E}_3$ for the DSBS. $C_3(D_1, D_2) = C(X_1, X_2)$ in the shaded region.

The joint rate distortion function of the DSBS (X_1, X_2) is given by [30]

$$R_{X_1X_2}(D_1, D_2) = \begin{cases} 1 + h(a_0) - h(D_1) - h(D_2), & (D_1, D_2) \in \mathcal{E}_1, \\ 1 - (1-a_0)h\left(\frac{D_1+D_2-a_0}{2(1-a_0)}\right) - a_0 h\left(\frac{D_1-D_2+a_0}{2a_0}\right), & (D_1, D_2) \in \mathcal{E}_2, \\ 1 - h\left(\min\{D_1, D_2\}\right), & (D_1, D_2) \in \mathcal{E}_3. \end{cases} \quad (56)$$

where $\mathcal{E}_1 = \mathcal{E}_{10} \cup \mathcal{E}_{11}$ with $\mathcal{E}_{10}, \mathcal{E}_{11}, \mathcal{E}_2$ and $\mathcal{E}_3$ defined in (55). Therefore, for this DSBS, $R_{X_1}(D_1) + R_{X_2}(D_2) - I(X_1; X_2) = R_{X_1X_2}(D_1, D_2)$, for $(D_1, D_2) \in \mathcal{E}_1$. From Lemma 5, we have for $(D_1, D_2) \in \mathcal{E}_1$,

$$C_3(D_1, D_2) \geq C(X_1, X_2). \quad (57)$$

On the other hand, the conditional rate distortion function $R_{X_i|S}(D_i)$, $i = 1, 2$, is given by [19]

$$R_{X_i|S}(D_i) = \begin{cases} h(a_1) - h(D_i), & 0 \leq D_i \leq a_1, \\ 0, & D_i \geq a_1. \end{cases}$$

Therefore, $R_{X_1|S}(D_1) + R_{X_2|S}(D_2) + I(X_1, X_2; S) = R_{X_1X_2}(D_1, D_2)$ is satisfied for $(D_1, D_2) \in \mathcal{E}_{10}$. By Theorem 5, $C_3(D_1, D_2) \leq C(X_1, X_2)$ for $(D_1, D_2) \in \mathcal{E}_{10}$. Together with (57) and given that $\mathcal{E}_{10} \subset \mathcal{E}_1$, we have proved that for $(D_1, D_2) \in \mathcal{E}_{10}$,

$$C_3(D_1, D_2) = C(X_1, X_2).$$

For $(D_1, D_2) \in \mathcal{E}_2$, we only need to show that $C_3(D_1, D_2) \geq R_{X_1X_2}(D_1, D_2)$. It was shown in [30] that the backward test channel that achieves $R_{X_1X_2}(D_1, D_2)$ is given by

$$\begin{aligned} X_1 &= \hat{X}_1 + Z_1, \\ X_2 &= \hat{X}_2 + Z_2, \end{aligned}$$

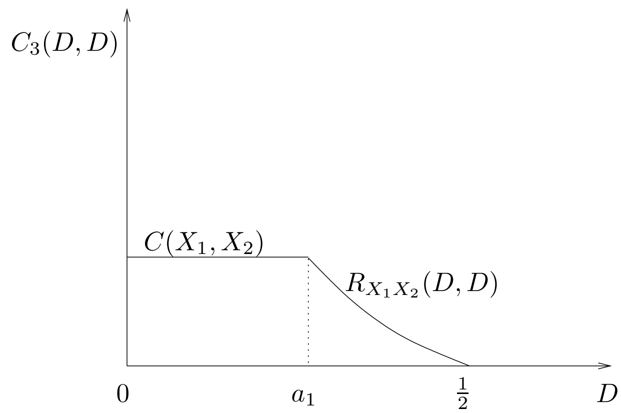

Fig. 5. The relation of $C_3(D,D)$ and D for the DSBS with $D_1 = D_2 = D$.

where both $\hat{X}_1, \hat{X}_2$ and Z_1, Z_2 are binary vectors independent of each other with the probability mass functions given respectively as

$$P_{\hat{X}_1\hat{X}_2} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}, \qquad P_{Z_1Z_2} = \frac{1}{2}\begin{bmatrix} 2-a_0-D_1-D_2 & D_2-D_1+a_0 \\ D_1-D_2+a_0 & D_1+D_2-a_0 \end{bmatrix}.$$

Therefore, $(\hat{X}_1, \hat{X}_2)$ that achieves $R_{X_1X_2}(D_1, D_2)$ satisfies

$$\hat{X}_2 = \hat{X}_1.$$

For the characterization $C^*(D_1, D_2)$ of $C_3(D_1, D_2)$, any W satisfying the Markov chain $\hat{X}_1 - W - \hat{X}_1$ must satisfy $H(\hat{X}_1|W) = 0$. Thus, $\hat{X}_1$ is a function of W and we have

$$I(X_1, X_2; W) = I(X_1, X_2; W, \hat{X}_1) \geq I(X_1, X_2; \hat{X}_1) = R_{X_1X_2}(D_1, D_2).$$

Therefore, $C_3(D_1, D_2) = R_{X_1X_2}(D_1, D_2)$.

The region $\mathcal{E}_3$ is a degenerated one. For example, $R_{X_1X_2}(D_1, D_2) = R_{X_1}(D_1)$ if $a_0 < \frac{D_2-D_1}{1-2D_1}$ and $D_i \leq \frac{1}{2}, i = 1, 2$. This implies that the optimal coding scheme is to ignore X_2 and optimally compress X_1. Then $\hat{X}_2$ can be estimated from $\hat{X}_1$ with distortion less than D_2. The case of $a_0 < \frac{D_1-D_2}{1-2D_2}$ is dealt with similarly. Hence, similar to the region $\mathcal{E}_2$, $C_3(D_1, D_2) = R_{X_1X_2}(D_1, D_2)$. ■

The characterization of $C_3(D_1, D_2)$ is plotted in Fig. 4 as a function of the distortion constraints. $C_3(D_1, D_2) = C(X_1, X_2)$ in the shaded region. For the symmetric distortion constraint, $D_1 = D_2 = D$, the relation of $C_3(D, D)$ and D for the DSBS is given in Fig. 5.

Remarks:

- The claim $C_3(D_1, D_2) = C(X_1, X_2)$ for $(D_1, D_2) \in \mathcal{E}_{10}$ can also be proved using Theorem 7. $R_{X_1X_2}(a_1, a_1)$ is achieved by the backward test channel $p_b(x_1, x_2|s) = p(x_1|s)p(x_2|s)$. The vector source (X_1, X_2) is successively refinable for any $(D_1, D_2) \leq (a_1, a_1)$ [30] and the scalar source X_i is successively refinable for any $D_i \leq a_1$, $i = 1, 2$ [27]. Thus by Theorem 7, $C_3(D_1, D_2) = C(X_1, X_2)$ for $(D_1, D_2) \leq (a_1, a_1)$.

- We have the full characterization of $C_3(D_1, D_2)$ in the distortion region except the region $\mathcal{E}_{11}$. From the proof of Proposition 2, we know that $C_3(D_1, D_2) \geq C(X_1, X_2)$ for $(D_1, D_2) \in \mathcal{E}_{11}$, but the exact value of $C_3(D_1, D_2)$ in this region remains unknown.
- Let $(D_1, D_2) \leq (D_1', D_2') \leq (a_1, a_1)$, then the rate $R_{X_1X_2}(D_1', D_2')$ is $(D_1, D_2)-$achievable in the Gray-Wyner network, i.e., $R_{X_1X_2}(D_1', D_2') \geq C_3(D_1, D_2)$.

  To show this, let $(\hat{X}_1, \hat{X}_2)$ achieve $R_{X_1X_2}(D_1', D_2')$. The backward test channel that achieves $R_{X_1X_2}(D_1', D_2')$ satisfies $p_b(x_1, x_2|\hat{x}_1\hat{x}_2) = p_b(x_1|\hat{x}_1)p_b(x_2|\hat{x}_2)$ where
$$p_b(x_i|\hat{x}_i) = \begin{cases} 1 - D_i', & \text{if } x_i = \hat{x}_i, \\ D_i', & \text{Otherwise.} \end{cases}$$
  for $i = 1, 2$. Then for $(D_1, D_2) \leq (D_1', D_2') \leq (a_1, a_1)$, let the rate allocation of R_0, R_1, R_2 in the Gray-Wyner network be
$$\begin{aligned} R_0 &= R_{X_1X_2}(D_1', D_2') = 1 + h(a_0) - h(D_1') - h(D_2'), \\ R_i &= R_{X_i|\hat{X}_1\hat{X}_2}(D_i) = R_{X_i|\hat{X}_i}(D_i) = h(D_i') - h(D_i), i = 1, 2. \end{aligned} \tag{58}$$
  Since R_0, R_1 and R_2 in (58) sum up to $R_{X_1X_2}(D_1, D_2)$, $R_{X_1X_2}(D_1', D_2')$ is $(D_1, D_2)-$achievable. The minimal R_0 satisfying (58) is exactly $C(X_1, X_2)$, which is achieved by letting $(D_1', D_2') = (a_1, a_1)$.

B. *Gaussian random variables*

In this section we consider bivariate Gaussian random variables X_1, X_2 with zero mean and covariance matrix
$$K_2 = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}. \tag{59}$$
The common information between this pair of Gaussian random variables is given in the following theorem.

Theorem 8: For two joint Gaussian random variables X_1, X_2 with covariance matrix K_2, the common information is
$$C(X_1, X_2) = \frac{1}{2}\log\frac{1+\rho}{1-\rho}. \tag{60}$$

Proof: See Appendix F. ■

As the common information of (X_1, X_2) is only a function of the correlation coefficient ρ, we consider, without loss of generality, the covariance matrix
$$K_2' = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \tag{61}$$

The above result generalizes to multi-variate Gaussian random variables satisfying a certain covariance matrix structure, the proof of which can be constructed in a similar fashion.

Corollary 1: For N joint Gaussian random variables $X_1, X_2, \cdots, X_N$ with covariance matrix K_N,

$$K_N = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \cdot & \cdot & \cdots & \cdot \\ \rho & \rho & \cdots & 1 \end{bmatrix}, \tag{62}$$

the common information is

$$C(\mathbf{X}_N) = \frac{1}{2}\log\left(1 + \frac{N\rho}{1-\rho}\right). \tag{63}$$

We now characterize the minimum common rate $C_3(D_1, D_2)$ in the Gray-Wyner lossy source coding network for bivariate Gaussian random variables with covariance matrix K_2' in equation (61). It was shown in [17] that for symmetric distortion, i.e.,$D_1 = D_2 = D$,

$$C_3(D, D) = \begin{cases} C(X_1, X_2), & 0 \le D \le 1-\rho, \\ R_{X_1X_2}(D, D), & 1-\rho \le D \le 1, \\ 0, & D \ge 1. \end{cases} \tag{64}$$

We characterize $C_3(D_1, D_2)$ for general distortion (D_1, D_2) in the following proposition.

Proposition 3: For bivariate Gaussian random variables X_1, X_2 with zero mean, covariance matrix K_2' and squared error distortion, we have that

$$C_3(D_1, D_2) = \begin{cases} C(X_1, X_2), & (D_1, D_2) \in \mathcal{D}_{10}, \\ R_{X_1X_2}(D_1, D_2), & (D_1, D_2) \in \mathcal{D}_2 \cup \mathcal{D}_3, \\ 0, & (D_1, D_2) \ge (1, 1), \end{cases} \tag{65}$$

$$C(X_1, X_2) \le C_3(D_1, D_2) \le R_{X_1X_2}(D_1, D_2), \quad (D_1, D_2) \in \mathcal{D}_{11}, \tag{66}$$

where

$$\begin{aligned} \mathcal{D}_{10} &= \{(D_1, D_2) : 0 \le D_i \le 1-\rho, i = 1, 2\}, \\ \mathcal{D}_{11} &= \mathcal{D}_{10}^c \cap \{(D_1, D_2) : D_1 + D_2 - D_1D_2 \le 1-\rho^2\}, \\ \mathcal{D}_2 &= \mathcal{D}_{10}^c \cap \mathcal{D}_{11}^c \cap \left\{(D_1, D_2) : \min\left\{\tfrac{1-D_1}{1-D_2}, \tfrac{1-D_2}{1-D_1}\right\} \ge \rho^2\right\}, \\ \mathcal{D}_3 &= \mathcal{D}_{10}^c \cap \mathcal{D}_{11}^c \cap \mathcal{D}_2^c \cap \{(D_1, D_2) : D_i \le 1, i = 1, 2\}. \end{aligned} \tag{67}$$

Proof: The joint rate distortion function for Gaussian random variables with squared error distortion [28]–[30] is given by

$$R_{X_1X_2}(D_1, D_2) = \begin{cases} \frac{1}{2}\log\frac{1-\rho^2}{D_1D_2}, & (D_1, D_2) \in \mathcal{D}_1, \\ \frac{1}{2}\log\frac{1-\rho^2}{D_1D_2 - \left(\rho - \sqrt{(1-D_1)(1-D_2)}\right)^2}, & (D_1, D_2) \in \mathcal{D}_2, \\ \frac{1}{2}\log\frac{1}{\min\{D_1, D_2\}}, & (D_1, D_2) \in \mathcal{D}_3, \end{cases} \tag{68}$$

where $\mathcal{D}_1 = \mathcal{D}_{10} \cup \mathcal{D}_{11}$. The marginal rate distortion function for $X_i \sim \mathcal{N}(0, 1), i = 1, 2$, is

$$R_{X_i}(D_i) = \begin{cases} \frac{1}{2}\log\frac{1}{D_i}, & 0 \le D_i \le 1, \\ 0, & D_i \ge 1. \end{cases}$$

Therefore, $R_{X_1}(D_1) + R_{X_2}(D_2) - I(X_1;X_2) = R_{X_1X_2}(D_1,D_2)$, for $(D_1,D_2) \in \mathcal{D}_1$. From Lemma 5, for $(D_1,D_2) \in \mathcal{D}_1$,

$$C_3(D_1,D_2) \geq C(X_1,X_2).$$

On the other hand, the random variable W in the following decomposition of X_1 and X_2 achieves the common information

$$X_i = \sqrt{\rho}W + \sqrt{1-\rho}N_i,\ i = 1,2. \tag{69}$$

where W, N_1, N_2 are mutually independent standard Gaussian random variables. The conditional distribution of X given W is Gaussian distribution with variance $1-\rho$. Hence, for $i = 1,2$, the conditional rate distortion function is

$$R_{X_i|W}(D_i) = \begin{cases} \frac{1}{2}\log\frac{1-\rho}{D_i}, & 0 \leq D_i \leq 1-\rho, \\ 0, & D_i \geq 1-\rho. \end{cases} \tag{70}$$

The condition $R_{X_1|W}(D_1)+R_{X_2|W}(D_2)+I(X_1,X_2;W) = R_{X_1X_2}(D_1,D_2)$ is satisfied for $(D_1,D_2) \in \mathcal{D}_{10}$. From Theorem 5, $C_3(D_1,D_2) \leq C(X_1,X_2)$ for $(D_1,D_2) \in \mathcal{D}_{10}$. Since, $\mathcal{D}_{10} \in \mathcal{D}_1$, we proved that for $(D_1,D_2) \in \mathcal{D}_{10}$,

$$C_3(D_1,D_2) = C(X_1,X_2).$$

For $(D_1,D_2) \in \mathcal{D}_2$, it was shown in [30] that $(\hat{X}_1,\hat{X}_2)$ that achieves $R_{X_1X_2}(D_1,D_2)$ satisfies

$$\hat{X}_2 = \sqrt{\frac{1-D_2}{1-D_1}}\hat{X}_1.$$

Hence, using the characterization $C^*(D_1,D_2)$, it is easy to show that the W satisfying the Markov chains (38) and (39) must satisfy two Markov chains

$$X_1X_2 - \hat{X}_1 - W - \hat{X}_2,$$
$$X_1X_2 - \hat{X}_2 - W - \hat{X}_1.$$

Therefore, we have

$$I(X_1,X_2;W) = I(X_1,X_2;\hat{X}_1) = I(X_1,X_2;\hat{X}_1,\hat{X}_2),$$

which proved $C_3(D_1,D_2) = R_{X_1X_2}(D_1,D_2)$.

The region $\mathcal{D}_3$ is a degenerated one. For example, $R_{X_1X_2}(D_1,D_2) = R_{X_1}(D_1)$ if $\frac{1-D_2}{1-D_1} < \rho^2$, this means that the correlation between X_1 and X_2 is so strong that the optimal coding scheme is to encode X_1 to within distortion D_1 and ignore X_2. Then $\hat{X}_2$ can be estimated from $\hat{X}_1$. We have

$$\hat{X}_2 = \rho\hat{X}_1.$$

The case of $\frac{1-D_1}{1-D_2} < \rho^2$ is dealt with similarly. Hence, we have $C_3(D_1,D_2) = R_{X_1X_2}(D_1,D_2)$. ■

The characterization of $C_3(D_1,D_2)$ is plotted in Fig. 6 as a function of the distortion constraints. $C_3(D_1,D_2) = C(X_1,X_2)$ in the shaded region.

Remarks:

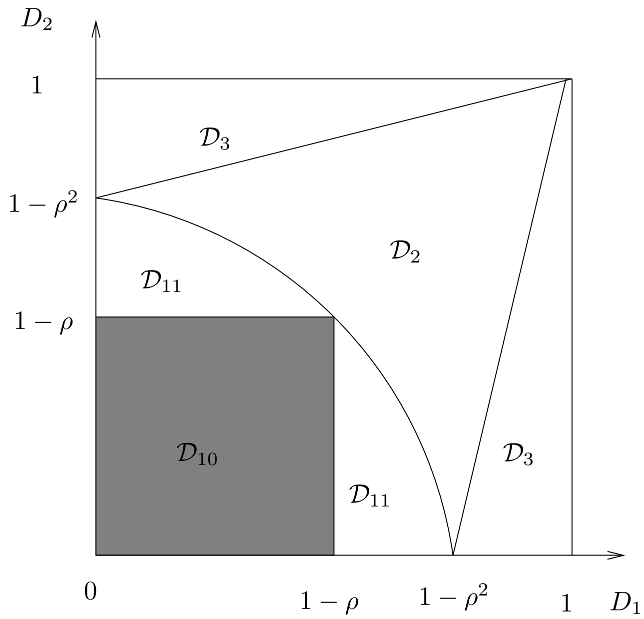

Fig. 6. The distortion regions $\mathcal{D}_{10}, \mathcal{D}_{11}, \mathcal{D}_2$ and $\mathcal{D}_3$ for bivariate Gaussian random variables. $C_3(D_1, D_2) = C(X_1, X_2)$ in the shaded region.

- Similar to the binary case, the claim $C_3(D_1, D_2) = C(X_1, X_2)$ for $(D_1, D_2) \in \mathcal{D}_{10}$ can also be proved using Theorem 7. This is because for the bivariate Gaussian random variables with covariance matrix K_2', $R_{X_1X_2}(1-\rho, 1-\rho)$ is achieved by the backward test channel $p_b(x_1, x_2|w) = p(x_1|w)p(x_2|w)$, (X_1, X_2) is successively refinable for any $(D_1, D_2) \leq (1-\rho, 1-\rho)$ [30] and X_i is successively refinable for $D_i \leq 1-\rho$, $i = 1, 2$ [27].
- Similarly, $C_3(D_1, D_2) \geq C(X_1, X_2)$ for $(D_1, D_2) \in \mathcal{D}_{11}$ but the exact characterization of $C_3(D_1, D_2)$ remains unknown in this region.
- Let $(D_1, D_2) \leq (D_1', D_2') \leq (1-\rho, 1-\rho)$, then the rate $R_{X_1X_2}(D_1', D_2')$ is $(D_1, D_2)-$achievable in the Gray-Wyner network, i.e., $R_{X_1X_2}(D_1', D_2') \geq C_3(D_1, D_2)$.

  This is because for $(D_1', D_2') \in \mathcal{E}_{10}$, the joint rate distortion function $R_{X_1X_2}(D_1', D_2')$ is achieved by Gaussian distributed $(\hat{X}_1, \hat{X}_2)$ satisfying $X_1 - \hat{X}_1 - \hat{X}_2 - X_2$ where the covariance matrix of $(\hat{X}_1, \hat{X}_2)$ is [30]

$$K_{\hat{X}_1\hat{X}_2} = \begin{bmatrix} 1-D_1' & \rho \\ \rho & 1-D_2' \end{bmatrix}.$$

  Then for $(D_1, D_2) \leq (D_1', D_2') \leq (1-\rho, 1-\rho)$, let the rate allocation of R_0, R_1, R_2 for the Gray-Wyner network be as follows:

$$\begin{aligned} R_0 &= R_{X_1X_2}(D_1', D_2') = \tfrac{1}{2}\log\tfrac{1-\rho^2}{D_1'D_2'}, \\ R_i &= R_{X_i|\hat{X}_1\hat{X}_2}(D_i) = R_{X_i|\hat{X}_i}(D_i) = \tfrac{1}{2}\log\tfrac{D_i'}{D_i}, i = 1, 2. \end{aligned} \tag{71}$$

  R_0, R_1 and R_2 in (71) sum up to $R_{X_1X_2}(D_1, D_2)$, so $R_{X_1X_2}(D_1', D_2')$ is $(D_1, D_2)-$achievable.

Therefore, in the Gray-Wyner network, we can use the rate allocation in (71) to achieve the distortion $(D_1, D_2) \leq (1-\rho, 1-\rho)$ for any $(D_1, D_2) \leq (D_1', D_2') \leq (1-\rho, 1-\rho)$. The minimal R_0 satisfying (71) is exactly $C(X_1, X_2)$, which is achieved by letting $(D_1', D_2') = (1-\rho, 1-\rho)$.

VI. Conclusion

We have generalized the definition of Wnyer's common information and expanded its practical significance by providing a new operational interpretation. The generalization is two-folded: the number of dependent variables can be arbitrary, so are the alphabet of those random variables. We have determined new properties for the generalized Wyner's common information of N dependent variables. More importantly, we have derived a lossy source coding interpretation of Wyner's common information using the Gray-Wyner network. In particular, it is established that the common information is precisely the smallest common message rate when the total rate is arbitrarily close to the rate distortion function with joint decoding. A surprising observation is that such equality holds independent of the values of distortion constraints as long as the distortions are within some distortion region. Two examples, the doubly symmetric binary source under Hamming distortion and bivariate Gaussian source under square-error distortion, are used to illustrate the lossy source coding interpretation of Wyner's common information. The common information for bivariate Gaussian source and its extension to the multi-variate case has also been computed explicitly.

While the lossy source coding interpretation of Wyner's common information presented in this paper is limited to $N = 2$ random variables, the results can be extended to arbitrary N random variables.

Acknowledgment

The authors gratefully acknowledges Professor Paul Cuff of Princeton University for pointing out an error in an earlier proof of Theorem 8 given in [33].

Appendix

A. *Proof of Theorem 5*

We first show that $C_3(D_1, D_2) \geq \tilde{C}(D_1, D_2)$. Let R_0 be (D_1, D_2)-achievable, then there exists an (n, M_0, M_1, M_2) code such that (32)-(34) are satisfied. Define $R_i = \frac{1}{n} \log M_i$ for $i = 1, 2$. Since (R_0, R_1, R_2) is (D_1, D_2)-achievable, from Theorem 2, there exists a W such that

$$\begin{aligned} R_0 &\geq I(X_1, X_2; W), \\ R_i &\geq R_{X_i|W}(D_i), \qquad i = 1, 2 \end{aligned}$$

and for any $\epsilon > 0$,

$$\sum_{i=0}^{2} R_i \leq R_{X_1 X_2}(D_1, D_2) + \epsilon. \tag{72}$$

Therefore,

$$\begin{aligned} R_{X_1X_2}(D_1,D_2)+\epsilon &\geq \sum_{i=0}^{2} R_i \\ &\geq I(X_1,X_2;W)+\sum_{i=1}^{2} R_{X_i|W}(D_i) \\ &\geq I(X_1,X_2;W)+R_{X_1X_2|W}(D_1,D_2) \qquad (73) \\ &\geq R_{X_1X_2}(D_1,D_2) \qquad (74) \end{aligned}$$

where (73) is from (15b) and (74) comes from (14b). Thus, we have

$$I(X_1,X_2;W)+R_{X_1|W}(D_1)+R_{X_2|W}(D_2)=R_{X_1X_2}(D_1,D_2). \qquad (75)$$

Hence, if R_0 is (D_1,D_2)-achievable, there exists a W such that $R_0 \geq I(X_1,X_2;W)$ and (75) is true. It shows that $C_3(D_1,D_2) \geq \tilde{C}(D_1,D_2)$.

Next we show $C_3(D_1,D_2) \leq \tilde{C}(D_1,D_2)$. Let W' be the random variable that achieves $\tilde{C}(D_1,D_2)$. For any $R_0 > \tilde{C}(D_1,D_2)$ and $\epsilon > 0$, let

$$\epsilon_1 = \min\left\{\frac{\epsilon}{3}, R_0-\tilde{C}(D_1,D_2)\right\}, \qquad (76)$$

and hence $\epsilon_1 > 0$. From theorem 2, there exists an (n,M_0,M_1,M_2) code with $Ed_1(X_1,\hat{X}_1) \leq D_1$, $Ed_2(X_2,\hat{X}_2) \leq D_2$, and

$$\begin{aligned} \frac{1}{n}\log M_0 &\leq I(X_1,X_2;W')+\epsilon_1=\tilde{C}(D_1,D_2)+\epsilon_1 \leq R_0, \qquad (77) \\ \frac{1}{n}\log M_i &\leq R_{X_i|W'}(D_i)+\epsilon_1, \qquad (78) \end{aligned}$$

for $i=1,2$. Sum over (77) and (78), we get

$$\begin{aligned} \sum_{i=0}^{2}\frac{1}{n}\log M_i &\leq I(X_1,X_2;W')+\sum_{i=1}^{2} R_{X_i|W'}(D_i)+3\epsilon_1 \\ &\leq R_{X_1X_2}(D_1,D_2)+\epsilon, \qquad (79) \end{aligned}$$

where inequality (79) comes from (76) and definition of $\tilde{C}(D_1,D_2)$.

This proves that R_0 is (D_1,D_2)-achievable, thus completes the proof of $C_3(D_1,D_2) \leq \tilde{C}(D_1,D_2)$.

B. *Direct proof of* $\tilde{C}(D_1,D_2)=C^*(D_1,D_2)$

First we show that $\tilde{C}(D_1,D_2) \geq C^*(D_1,D_2)$.

Let W be the variable that achieves $\tilde{C}(D_1,D_2)$ and let $\hat{X}_1,\hat{X}_2$ be random variables that achieve $R_{X_1|W}(D_1)$

and $R_{X_2|W}(D_2)$, i.e.,

$$\begin{aligned}
I(X_1, X_2; W) + R_{X_1|W}(D_1) + R_{X_2|W}(D_2) &= R_{X_1X_2}(D_1, D_2), && (80)\\
R_{X_1|W}(D_1) &= I(X_1; \hat{X}_1|W), && (81)\\
R_{X_2|W}(D_2) &= I(X_2; \hat{X}_2|W), && (82)\\
E[d_1(X_1, \hat{X}_1)] &\le D_1, && (83)\\
E[d_2(X_2, \hat{X}_2)] &\le D_2. && (84)
\end{aligned}$$

Without loss of generality, we can assume that the joint distribution of $(X_1, X_2, \hat{X}_1, \hat{X}_2, W)$ factors as $p(x_1, x_2, \hat{x}_1, \hat{x}_2, w) = p(x_1, x_2, w)p(\hat{x}|x, w)p(\hat{y}|y, w)$ because the distortion D_1 is independent of X_2 and D_2 is independent of X_1. We now establish

$$R_{X_1X_2|W}(D_1, D_2) = R_{X_1|W}(D_1) + R_{X_2|W}(D_2).$$

This is from (80) and the inequalities

$$\begin{aligned}
R_{X_1X_2|W}(D_1, D_2) + I(X_1, X_2; W) &\ge R_{X_1X_2}(D_1, D_2),\\
R_{X_1|W}(D_1) + R_{X_2|W}(D_2) &\ge R_{X_1X_2|W}(D_1, D_2),
\end{aligned}$$

from Lemma 1. Therefore, together with (80)-(84), we have

$$\begin{aligned}
R_{X_1X_2|W}(D_1, D_2) &= I(X_1; \hat{X}_1|W) + I(X_2; \hat{X}_2|W)\\
&= H(\hat{X}_1|W) + H(\hat{X}_2|W) - H(\hat{X}_1|X_1, W) - H(\hat{X}_2|X_2, W)\\
&\ge H(\hat{X}_1, \hat{X}_2|W) - H(\hat{X}_1|X_1, W) - H(\hat{X}_2|X_2, W)\\
&= H(\hat{X}_1, \hat{X}_2|W) - H(\hat{X}_1|W, X_1, X_2) - H(\hat{X}_2|W, X_1, X_2)\\
&= I(X_1, X_2; \hat{X}_1, \hat{X}_2|W)\\
&\ge R_{X_1X_2|W}(D_1, D_2).
\end{aligned}$$

As the left-hand side (LHS) and right-hand side (RHS) of the above inequalities are the same, all the inequalities must be equalities so we have

$$I(\hat{X}_1; \hat{X}_2|W) = 0.$$

Then we have

$$\begin{aligned}
R_{X_1X_2}(D_1, D_2) &= I(X_1, X_2; W) + I(X_1; \hat{X}_1|W) + I(X_2; \hat{X}_2|W)\\
&= I(X_1, X_2; W, \hat{X}_1, \hat{X}_2) - I(X_1, X_2; \hat{X}_1, \hat{X}_2|W) + I(X_1; \hat{X}_1|W) + I(X_2; \hat{X}_2|W)\\
&= I(X_1, X_2; \hat{X}_1, \hat{X}_2) + I(X_1, X_2; W|\hat{X}_1, \hat{X}_2)\\
&\ge I(X_1, X_2; \hat{X}_1, \hat{X}_2)\\
&\ge R_{X_1X_2}(D_1, D_2).
\end{aligned}$$

As the LHS and RHS of the above inequalities are the same, all the inequalities must be equalities so we have

$$\begin{aligned} I(X_1, X_2; W|\hat{X}_1, \hat{X}_2) &= 0, \\ I(X_1, X_2; \hat{X}_1, \hat{X}_2) &= R_{X_1X_2}(D_1, D_2). \end{aligned}$$

Therefore, $X_1, X_2, \hat{X}_1, \hat{X}_2, W$ satisfy the Markov chains in (38) and (39) and $\hat{X}_1, \hat{X}_2$ achieve $R_{X_1X_2}(D_1, D_2)$. Thus, $\tilde{C}(D_1, D_2) \geq C^*(D_1, D_2)$.

Next we show that $\tilde{C}(D_1, D_2) \leq C^*(D_1, D_2)$.

Let $X_1, X_2, X_1^*, X_2^*, W$ achieve $C^*(D_1, D_2)$. Therefore, they satisfy the Markov chains in (38) and (39) and $I(X_1, X_2; X_1^*, X_2^*) = R_{X_1X_2}(D_1, D_2)$ and $E[d_1(X_1, X_1^*)] \leq D_1, E[d_2(X_2, X_2^*)] \leq D_2$.

$$\begin{aligned} R_{X_1X_2}(D_1, D_2) &= I(X_1, X_2; X_1^*, X_2^*) \\ &= I(X_1, X_2; W, X_1^*, X_2^*) && (85) \\ &= I(X_1, X_2; W) + I(X_1, X_2; X_1^*, X_2^*|W) \\ &= I(X_1, X_2; W) + H(X_1^*|W) + H(X_2^*|W) - H(X_1^*, X_2^*|X_1, X_2, W) && (86) \\ &= I(X_1, X_2; W) + I(X_1; X_1^*|W) + I(X_2; X_2^*|W) + H(X_1^*|X_1, W) \\ &\quad + H(X_2^*|X_2, W) - H(X_1^*, X_2^*|X_1, X_2, W) \\ &\geq I(X_1, X_2; W) + I(X_1; X_1^*|W) + I(X_2; X_2^*|W) + H(X_1^*|X_1, X_2, W) \\ &\quad + H(X_2^*|X_1, X_2, W) - H(X_1^*, X_2^*|X_1, X_2, W) && (87) \\ &= I(X_1, X_2; W) + I(X_1; X_1^*|W) + I(X_2; X_2^*|W) + I(X_1^*; X_2^*|X_1, X_2, W) \\ &\geq I(X_1, X_2; W) + I(X_1; X_1^*|W) + I(X_2; X_2^*|W) \\ &\geq I(X_1, X_2; W) + R_{X_1|W}(D_1) + R_{X_2|W}(D_2) \\ &\geq I(X_1, X_2; W) + R_{X_1X_2|W}(D_1, D_2) && (88) \\ &\geq R_{X_1X_2}(D_1, D_2), && (89) \end{aligned}$$

where (85) is from the Markov chain $(X_1, X_2) - (X_1^*, X_2^*) - W$, (86) is from the Markov chain $X_1^* - W - X_2^*$, (87) is because conditioning reduces entropy, (88) and (89) are by the properties of rate distortion functions. As the LHS and RHS of the above inequalities are the same, all the inequalities must be equalities so we have

$$I(X_1, X_2; W) + R_{X_1|W}(D_1) + R_{X_2|W}(D_2) = R_{X_1X_2}(D_1, D_2).$$

Therefore, $C^*(D_1, D_2) = I(X_1, X_2; W) \geq \tilde{C}(D_1, D_2)$.

C. *Proof of Lemma 5*

Let W be the random variable that achieves $C_3(D_1, D_2)$. Thus, $C_3(D_1, D_2) = I(X_1, X_2; W)$ with

$$R_{X_1|W}(D_1) + R_{X_2|W}(D_2) + I(X_1, X_2; W) = R_{X_1X_2}(D_1, D_2). \tag{90}$$

Combined with (42), we have that

$$\begin{aligned} R_{X_1}(D_1) + R_{X_2}(D_2) - I(X_1;X_2) &= R_{X_1|W}(D_1) + R_{X_2|W}(D_2) + I(X_1,X_2;W) && (91)\\ &\geq R_{X_1}(D_1) - I(X_1;W) + R_{X_2}(D_2) - I(X_2;W) \\ &\quad + I(X_1,X_2;W) && (92)\\ &= R_{X_1}(D_1) + R_{X_2}(D_2) - I(X_1;X_2) + I(X_1;X_2|W) && (93)\\ &\geq R_{X_1}(D_1) + R_{X_2}(D_2) - I(X_1;X_2), && (94) \end{aligned}$$

where equation (91) is from equations (90) and (42), inequality (92) comes from Lemma 1, (93) is by the chain rule and inequality (94) is by the fact that $I(X_1;X_2|W) \geq 0$.

Because the LHS of (91) is the same as the RHS of (94), we can conclude that all the inequalities above should be equalities. This implies $I(X_1;X_2|W) = 0$. Therefore,

$$C_3(D_1,D_2) \geq C(X_1,X_2).$$

D. Proof of Theorem 6

Let W be the random variable that achieves the common information of X_1, X_2. By Lemma 2, there exists a strictly positive surface $\mathcal{D}(X_1X_2|W)$ such that for any $0 \leq (D_1,D_2) \leq \mathcal{D}(X_1X_2|W)$,

$$I(X_1,X_2;W) + R_{X_1X_2|W}(D_1,D_2) = R_{X_1X_2}(D_1,D_2). \tag{95}$$

Also by Lemma 2, there exists a strictly positive surface $\mathcal{D}(X_1X_2) \geq \mathcal{D}(X_1X_2|W)$ such that for any $0 \leq (D_1,D_2) \leq \mathcal{D}(X_1X_2)$,

$$R_{X_1}(D_1) + R_{X_2}(D_2) - I(X_1;X_2) = R_{X_1X_2}(D_1,D_2). \tag{96}$$

Since $\mathcal{D}(X_1X_2|W) \leq \mathcal{D}(X_1X_2)$, let $\gamma = \mathcal{D}(X_1X_2|W)$, both equalities (95) and (96) hold for $0 \leq (D_1,D_2) \leq \gamma$. Therefore, from Lemmas 4 and 5, $C_3(D_1,D_2) = C(X_1,X_2)$ for $0 \leq (D_1,D_2) \leq \gamma$.

E. Proof of Theorem 7

First we show that for any $(D_1,D_2) \leq (D_1^0,D_2^0)$,

$$R_{X_1X_2|W}(D_1,D_2) + I(X_1X_2;W) = R_{X_1X_2}(D_1,D_2). \tag{97}$$

From the definition of (D_1^0,D_2^0) in (45), we have

$$R_{X_1X_2}(D_1^0,D_2^0) \geq R_{X_1}(D_1^0) + R_{X_2}(D_2^0) - I(X_1;X_2) = I(X_1,X_2;W),$$

where the first inequality is from (14c). On the other hand,

$$R_{X_1X_2}(D_1^0,D_2^0) \leq I(X_1,X_2;\hat{X}_1^0,\hat{X}_2^0) \leq I(X_1,X_2;W).$$

Therefore, $R_{X_1X_2}(D_1^0,D_2^0) = I(X_1,X_2;\hat{X}_1^0,\hat{X}_2^0) = I(X_1X_2;W)$.

Let $(\hat{X}_1, \hat{X}_2)$ achieve $R_{X_1X_2}(D_1, D_2)$. As the vector source (X_1, X_2) is successively refinable under individual distortion constraints [30], we have the Markov chain $X_1X_2 - \hat{X}_1\hat{X}_2 - \hat{X}_1^0\hat{X}_2^0$. Therefore,

$$\begin{aligned}
R_{X_1X_2}(D_1, D_2) - I(X_1, X_2; W) &= I(X_1, X_2; \hat{X}_1, \hat{X}_2) - I(X_1, X_2; \hat{X}_1^0, \hat{X}_2^0) \\
&= I(X_1X_2; \hat{X}_1\hat{X}_2 | \hat{X}_1^0, \hat{X}_2^0) \\
&\geq R_{X_1X_2|\hat{X}_1^0, \hat{X}_2^0}(D_1, D_2) \\
&\geq R_{X_1X_2|W}(D_1, D_2),
\end{aligned}$$

where the last inequality is from the Markov chain $X_1X_2 - W - \hat{X}_1^0, \hat{X}_2^0$. On the other hand, by Lemma 1, we have

$$R_{X_1X_2|W}(D_1, D_2) + I(X_1X_2; W) \geq R_{X_1X_2}(D_1, D_2).$$

This establishes (97). Thus, from Lemma 4, $C_3(D_1, D_2) \leq C(X_1; X_2)$.

To complete the proof, we need to show

$$R_{X_1}(D_1) + R_{X_2}(D_2) - I(X_1; X_2) = R_{X_1X_2}(D_1, D_2). \tag{98}$$

From Lemma 1,

$$R_{X_1}(D_1) + R_{X_2}(D_2) - I(X_1; X_2) \leq R_{X_1X_2}(D_1, D_2).$$

Therefore, we only need to establish the other direction. For $i = 1, 2$, let $\hat{X}_i$ achieve $R_{X_i}(D_i)$, then by the definition of a successively refinable scalar source [27], we have the Markov chain $X_i - \hat{X}_i - \hat{X}_i^0$ for $D_i \leq D_i^0$. Therefore,

$$\begin{aligned}
R_{X_i}(D_i) - I(X_i; W) &= I(X_i; \tilde{X}_i) - I(X_i; \hat{X}_i^0) \\
&= I(X_i; \hat{X}_i | \hat{X}_i^0) \\
&\geq R_{X_i|\hat{X}_i^0}(D_i) \\
&\geq R_{X_i|W}(D_i), \qquad (99)
\end{aligned}$$

where (99) is from the Markov chain $X_i - W - \hat{X}_i^0$. Using (99), we have

$$\begin{aligned}
R_{X_1}(D_1) + R_{X_2}(D_2) - I(X_1; X_2) &\geq R_{X_1|W}(D_1) + I(X_1; W) + R_{X_2|W}(D_1) + I(X_2; W) - I(X_1; X_2) \\
&= R_{X_1|W}(D_1) + R_{X_2|W}(D_2) + I(X_1X_2; W) \\
&= R_{X_1X_2|W}(D_1, D_2) + I(X_1X_2; W) \\
&= R_{X_1X_2}(D_1, D_2),
\end{aligned}$$

which completes the proof.

F. *Proof of Theorem 8*

First, we will show that the common information of X_1, X_2 is only a function of the correlation coefficient ρ. To show this, let $\tilde{X}_i = \frac{1}{\sigma_i} X_i,\ i = 1,2$, thus $\tilde{X}_1, \tilde{X}_2$ are joint Gaussian distributed with zero mean and covariance matrix

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

We have the Markov chain that $\tilde{X}_1 - X_1 - X_2 - \tilde{X}_2$ and by the data processing inequality for Wyner's common information [13], $C(\tilde{X}_1, \tilde{X}_2) \leq C(X_1, X_2)$. On the other hand, we have the Markov chain that $X_1 - \tilde{X}_1 - \tilde{X}_2 - X_2$ and $C(\tilde{X}_1, \tilde{X}_2) \leq C(X_1, X_2)$. Thus, $C(\tilde{X}_1, \tilde{X}_2) = C(X_1, X_2)$. Without loss generality, we will consider $\sigma_1^2 = \sigma_2^2 = 1$ in the following.

Let

$$X_i = \sqrt{\rho} W + \sqrt{1-\rho} N_i,\ i = 1,2, \tag{100}$$

where W, N_1, N_2 are mutually independent standard Gaussian random variables. It is clear that X_1, X_2 are bivariate Gaussian with correlation coefficient ρ,

$$C(X_1, X_2) \leq I(X_1, X_2; W) = \frac{1}{2} \log \frac{1+\rho}{1-\rho}.$$

Next we will show that

$$C(X_1, X_2) \geq \frac{1}{2} \log \frac{1+\rho}{1-\rho}.$$

For any U that satisfies the Markov chain $X_1 - U - X_2$, let D_1 be the minimum mean square error (MMSE) of estimating X_1 using U, thus, $D_1 = E(X_1 - E(X_1|U))^2$. Similarly, let $D_2 = E(X_2 - E(X_2|U))^2$. We now show that $I(X_1 X_2; U) \geq \frac{1}{2} \log \frac{1+\rho}{1-\rho}$.

$$\begin{aligned}
I(X_1 X_2; U) &= H(X_1 X_2) - H(X_1|U) - H(X_2|U) \\
&= I(X_1; U) + I(X_2; U) - I(X_1; X_2) && (101) \\
&\geq I(X_1; E(X_1|U)) + I(X_2; E(X_2|U)) - I(X_1; X_2) && (102) \\
&\geq R_{X_1}(D_1) + R_{X_2}(D_2) - I(X_1; X_2) && (103) \\
&= \frac{1}{2} \log \frac{1-\rho^2}{D_1 D_2},
\end{aligned}$$

for $D_1 \leq 1, D_2 \leq 1$, where (101) is from the chain rule, (102) is from the Markov chains $X_1 - U - E(X_1|U)$, $X_2 - U - E(X_2|U)$ and (103) is by the definition of rate distortion function.

Next we show that $D_1 + D_2 \leq 2(1-\rho)$, $D_1 \leq 1$, $D_2 \leq 1$.

$$\begin{aligned}
2(1-\rho) &= E(X_1 - X_2)^2 \\
&= E[X_1 - E(X_1|U) + E(X_1|U) - X_2]^2 \\
&= E[X_1 - E(X_1|U)]^2 + E[E(X_1|U) - X_2]^2 + 2E[(X_1 - E(X_1|U))(E(X_1|U) - X_2)] \\
&= E[X_1 - E(X_1|U)]^2 + E[E(X_1|U) - X_2]^2 \qquad (104) \\
&= E[X_1 - E(X_1|U)]^2 + E[E(X_1|U) - E(X_2|U) + E(X_2|U) - X_2]^2 \\
&= E[X_1 - E(X_1|U)]^2 + E[X_2 - E(X_2|U)]^2 + E[E(X_2|U) - E(X_1|U)]^2 \\
&\quad + E[(X_2 - E(X_2|U))(E(X_2|U) - E(X_1|U))] \\
&= E[X_1 - E(X_1|U)]^2 + E[X_2 - E(X_2|U)]^2 + E[E(X_2|U) - E(X_1|U)]^2 \qquad (105) \\
&\geq D_1 + D_2
\end{aligned}$$

where (104) is from

$$\begin{aligned}
E[(X_1 - E(X_1|U))(E(X_1|U) - X_2)] &= E[(X_1 - E(X_1|U))E(X_1|U)] - E[(X_1 - E(X_1|U))X_2] \\
&= -E[(X_1 - E(X_1|U))X_2] \\
&= -E_{UX_2}[X_2 E_{X_1|U}[X_1 - E(X_1|U)]] \\
&= -E_{UX_2}[X_2(E(X_1|U) - E(X_1|U))] = 0,
\end{aligned}$$

and (105) is from

$$\begin{aligned}
& E[(X_2 - E(X_2|U))(E(X_2|U) - E(X_1|U))] \\
= \; & E[(X_2 - E(X_2|U))E(X_2|U)] - E[(X_2 - E(X_2|U))E(X_1|U)] = 0
\end{aligned}$$

In addition, we have $D_1 = E[X_1 - E(X_1|U)]^2 = EX_1^2 - E[E(X_1|U)^2] \leq EX_1^2 = 1$.

Thus,

$$\begin{aligned}
I(X_1X_2;U) &\geq \frac{1}{2}\log\frac{1-\rho^2}{D_1D_2} \\
&\geq \frac{1}{2}\log\frac{1-\rho^2}{\left(\frac{D_1+D_2}{2}\right)^2} \\
&\geq \frac{1}{2}\log\frac{1-\rho^2}{(1-\rho)^2} \\
&= \frac{1}{2}\log\frac{1+\rho}{1-\rho}.
\end{aligned}$$

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379-423, 1948.

[2] P. Gács and J. Körner, "Common information is much less than mutual information," *Problems Contr. Inform. Theory*, vol. 2, pp. 149-162, 1973.

[3] R. Ahlswede and J. Körner, "On common information and related characteristics of correlated information sources," *Proc. of the 7th Prague Conference of Information Theory*, 1974.

[4] A. D. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 163-179, March, 1975.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

[6] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic, New York, 1981.

[7] R. Ahlswede and I. Csiszár, "Common randomness in information thoery and cryptography, Part I: Secret sharing," *IEEE Trans. Inf. Theory*, vol. 39, pp. 1121-1132, July 1993.

[8] R. Ahlswede and I. Csiszár, "Common randomness in information thoery and cryptography, Part II: CR capacity," *IEEE Trans. Inf. Theory*, vol. 44, pp. 225-240, Jan. 1998.

[9] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inf. Theory*, vol. 39, pp. 733-742, May 1993.

[10] S. Kamath, V. Anantharam, "A new dual to the Gács-Körner common information defined via the Gray-Wyner system," *Proc. 48th Annual Allerton Conference on Communications, Control and Computing*,Monticello, IL, Sept. 2010.

[11] H. S. Witsenhausen, "On sequences of pairs of dependent random varibles," *SIAM J. Appl. Math.*, vol. 28, pp. 100-113, Jan. 1975.

[12] P. Cuff, H.Permuter and T. M. Cover, "Coordination capacity," *IEEE Trans. Inf. Theory*, vol.56, pp. 4181-4206, Sept. 2010.

[13] H. S. Witsenhausen, "Values and bounds for the common information of two discrete random variables," *SIAM J. Appl. Math.*, vol. 31, no. 2, pp. 313-333, 1976.

[14] G. Hu "On the amount of information," *Teor. Veroyatnost. i Primenen.*, 4: 447-455, 1962 (in Russian).

[15] R. W. Yeung, *Information Theory and Network Coding*, Springer, 2008.

[16] H. Tyagi and P. Narayan and P. Gupta, "When is a Function Securely Computable? ," *IEEE Trans. Inf. Theory*, vol.57, pp. 6337-6350, Oct. 2011.

[17] K. Visvanatha, E. Akyol and K. Rose, "Lossy common information of two dependent random variables," *Proc. IEEE Symp. Inform. Thoery*, Combridge, MA, July, 2012.

[18] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *Bell Syst. Tech. J.*, vol. 58, pp. 1681-1721, Nov. 1974.

[19] R. M. Gray, "A new class of lower bounds to information rates of stationary sources via conditional rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 19, pp. 480-489, July 1973.

[20] B. M. Leiner, "An alternate proof of the composite lower bounds," *Information and Control*,vol. 33, pp. 72-86, 1977.

[21] R. M. Gray, "Conditional Rate-Distortion Theory," Stanford Electronic labs., Stanford, CA, Tech. Rep. (6502-2), Oct. 1972.

[22] G. Xu and B. Chen, "Information for inference," *Proc. 49th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sept. 2011.

[23] R. Venkataramani, G. Kramer and V. K. Goyal, "Multiple description coding with many channels," *IEEE Trans. Inf. Theory*, Vol. 49, no. 9, pp. 2106-2114, Sept. 2003.

[24] T. J. Pinkston, "An application of rate-distortion theory to a converse to the coding theorem," *IEEE Trans. Inf. Theory*, Vol. IT-15, pp. 66-71, 1969.

[25] F. Jelinek, "Evalution of distortion rate functions for low distortions," *Proc. IEEE*, Vol. 55, pp. 2067-2068, Nov. 1967.

[26] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder-II: General sources," *Information and control*, vol. 38, pp. 60-80, 1978.

[27] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 269-275, March 1991.

[28] T. Berger, *Rate Distortion Theory:A mathematical basis for data compression*, Englewood Cliffs, N.J.: Prentice-Hall, 1971.

[29] J.-J. Xiao and Z.-Q. Luo, "Compression of correlated Gaussian sources under individual distortion criteria," *Proc. 43rd Allerton Conf. Commun. Control Comput.*, pp. 438-447, Sept. 2005.

[30] J. Nayak, E. Tuncel, D. Gunduz and E. Erkip, "Successive refinement of vector sources under individual distortion criteria," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1769-1781, Apr. 2010.

[31] J. M. Bernardo, "The concept of Exchangeablity and its applications," *Far East Journal of Mathematical Sciences*, vol.4, pp. 111-121, 1996.

[32] W. Liu, G. Xu and B. Chen, “The common information of N dependent random variables,” *Proc. 48th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sept. 2010.

[33] G. Xu, W. Liu and B. Chen, “ Wyner’s common information for continuous random variables - a Lossy source coding interpretation,” *Proc. Annual Conference on Information Sciences and Systems* , Baltimore, MD, March 2011.

[34] R. Tandon, L.Sankar, and H. V. Poor, “Multi-user privacy: The Gray-Wyner system and generalized common information,” *Proc. IEEE Symp. Inform. Thoery*, St. Petersburg, Russia, Aug. 2011.